Supplementary materials for this article are available at https://doi.org/10.1007/s13253-024-00627-9.

CMPLE: Correlation Modeling to Decode Photosynthesis Using the Minorize–Maximize Algorithm

Abhijnan CHATTOPADHYAY, Donghee HOH, David M. KRAMER, Tapabrata MAITI, and Samiran SINHA

In plant genomic experiments, correlations among various biological traits (phenotypes) give new insights into how genetic diversity may have tuned biological processes to enhance fitness under diverse conditions. Consequently, knowing how the correlations are affected by genetic (G) and environmental (E) factors helps develop climate-resilient plants. However, the current literature lacks any method for assessing the effect of predictors on pairwise correlations among multiple phenotypes together with easily interpretable model parameters. To address this need, we propose to model pairwise correlations directly in terms of G and E and develop a computationally efficient inference procedure. Two major novelties in our methodology are (1) the use of a composite pairwise likelihood method to avoid the positive definiteness restriction on the correlation matrix and (2) the use of a novel Minorize–Maximize (MM) algorithm for the efficient estimation of a large number of parameters. The proposed method shows excellent numerical performance on synthetic datasets. The analysis of the motivating data on cowpea reveals that the rates of solar energy storage by photosynthesis (the aggregate trait) are differentially affected by different genetic loci through two distinct processes: "photoinhibition" which results from photodamage caused by excess light, and "photoprotection" which protects plants from photodamage but also results in energy loss.

Supplementary material to this paper is provided online.

Key Words: Correlation modeling; Gene–environment interaction; MM algorithm; Pairwise composite likelihood; Phenome-wide association study; Standard deviation modeling.

A. Chattopadhyay, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, Durham, NC, USA.
A. Chattopadhyay · D. Hoh · D. M. Kramer, MSU-DOE Plant Research Lab, East Lansing, MI, USA.
D. M. Kramer, Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, USA.
T. Maiti, Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA.
S. Sinha (✉), Department of Statistics, Texas A&M University, College Station, TX, USA (E-mail: *sinha@stat.tamu.edu*).

© *2024 International Biometric Society*
Journal of Agricultural, Biological, and Environmental Statistics
https://doi.org/10.1007/s13253-024-00627-9

Published online: 23 May 2024

1. INTRODUCTION

Identifying genetic variations that confer improved performance is key to breeding or engineering more resilient and productive crops. Several methods have been suggested for investigating the statistical association between measured traits and genetic markers, including genome-wide association studies (GWAS) and whole-genome regression (WGR) approaches, which produce familiar quantitative trait loci (QTL) maps (Gianola and Fernando 2020). Standard GWAS approaches have mainly been used to analyze the genetic association with individual traits, identifying the mean differences between the marker genotypes for a given trait. While most loci influence the mean value of a trait, alterations in certain genetic loci can affect the variability and correlations between multiple traits, which in turn can be classified as meaningful biological mechanisms (Hall et al. 2014). This is particularly important when addressing complex traits such as photosynthetic efficiency or crop yield, which are affected by multiple genetic (G) variants and can change under different environmental (E) conditions.

Our motivating example emerges from photosynthesis experiments (Kramer et al. 2004; Kanazawa et al. 2021). Biologically, plants cannot use all of their solar energy for photochemical reactions; the excess energy causes photodamage or photoprotection. This delicate balance of processes can be explored by studying the interrelations among the following traits, namely ϕ_{II}, NPQt, qL, and ECSt (more details on these traits are given in the data analysis section). It is expected that higher values of ϕ_{II} and lower values of NPQt create a genetically elite plant to tolerate different stress conditions (e.g., high heat); this implies a negative correlation between the two traits among these plants. Therefore, the knowledge of how the correlation is affected by G and E is necessary to optimize the genetics and management of crops. In turn, these optimal values of the estimated correlations from diverse species will reveal additional adaptation mechanisms and guide the breeding and engineering of photosynthesis for higher, more climate-resilient productivity.

However, no scalable statistical methods currently exist for conducting genome-wide G $\times$ E studies across multiple phenotypes within the scope of GWAS. While GWAS tools can pinpoint potential causal loci, they face limitations in detecting epistasis (genetic mutations influencing other mutations) and pleiotropy (single locus controlling multiple phenotypic characteristics). This gap between genetic studies and therapeutic applications drives research into the post-GWAS era, focusing on cross-phenotype association analysis, known as phenome-wide association study (PheWAS). However, genome-wide PheWAS analysis poses significant computational challenges due to analyzing high-dimensional data and estimating numerous parameters. Our method addresses these challenges by elucidating the functions of identified loci and understanding their contributions to complex trait pathogenesis through modeling pairwise correlations across phenotypes based on selected genes/markers. Applied in the context of photosynthesis, our approach aids plant breeders in selecting predictor variable ranges to achieve desirable values for multiple traits simultaneously.

In mathematical terms, let $\boldsymbol{Y}$ be a vector of q numerically observable traits. Let $\boldsymbol{X}$ denote a vector of p-predictors. Suppose that conditional on $\boldsymbol{X}$, $\boldsymbol{Y}$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu}(\boldsymbol{X})$ and $q \times q$ variance-covariance matrix $\boldsymbol{\Sigma}(X)$. The use

of $\boldsymbol{X}$ as the argument for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ signifies that these parameters are functions of $\boldsymbol{X}$. The literature on modeling the mean vector in terms of predictor is very rich. So, for the sake of simplicity of the method development, we set $\boldsymbol{\mu} = \mathbf{0}$. Let $\sigma_j(\boldsymbol{X})$ be the standard deviation of the jth trait. Then, the $(j,k)th$ element of $\boldsymbol{\Sigma}(\boldsymbol{X})$ is $\sigma_j(\boldsymbol{X})\sigma_k(\boldsymbol{X})\rho_{j,k}(\boldsymbol{X})$, where $\rho_{j,k}(\boldsymbol{X})$ denotes the correlation between the jth and the kth traits as a function of the predictors $\boldsymbol{X}$. We are interested in relating $\rho_{j,k}(\boldsymbol{X})$ to $\boldsymbol{X}$ for every (j,k) th pair.

Toward that goal, initially $\rho_{j,k}$ is expressed in terms of $\boldsymbol{X}$, that is, $g(\rho_{j,k}) = \delta_{j,k,0} + \boldsymbol{X}^\top \boldsymbol{\delta}_{j,k,1}$, for a suitable link function g. Also, write $h(\sigma_j) = \alpha_{j,0} + \boldsymbol{X}^\top \boldsymbol{\alpha}_{j,1}$, for an appropriate link function h. The estimation of these parameters is quite challenging. When $\boldsymbol{\Sigma}$ is expressed in terms of the free-floating parameters $\delta_{j,k,0}$, $\boldsymbol{\delta}_{j,k,1}$, $\alpha_{j,0}$ and $\boldsymbol{\alpha}_{j,1}$, there is no guarantee that $\boldsymbol{\Sigma}$ is positive definite (PD). There is no easy recipe for restricting the parameters to achieve positive definiteness of $\boldsymbol{\Sigma}$. This issue persists even when the number of predictors is one (i.e., $p = 1$). Although the literature on covariance estimation and modeling covariance in terms of predictors has been growing, to the best of our knowledge, there is little to no literature on both modeling covariances with regression techniques and estimating correlations in the multivariate setting, simultaneously. Before outlining our technique, we discuss related literature on this topic.

The literature review is divided into two main parts: modeling the covariance matrix and regression modeling of the covariance matrix. Pourahmadi (1999) introduced a modified Cholesky decomposition to express the covariance matrix, where the unconstrained entries are linked to predictors, ensuring positive definiteness. Similarly, Daniels and Pourahmadi (2002) utilized it to construct prior distributions for Bayesian estimation of covariance matrices, while Huang et al. (2006) proposed a modified version combined with penalized techniques to estimate large covariance matrices. However, when considering modeling the unrestricted entries of the Cholesky decomposition matrix in terms of predictor variables, it does not resolve our issue. This is because the resulting regression parameters lack straightforward interpretation, particularly when the scientific focus is on pairwise correlations between traits. To explain this more, suppose that there are two traits, Y_1 and Y_2, and X is a scalar covariate. Let $\Sigma(X)$ be the covariance between Y_1 and Y_2 given the covariate X. Let $\Sigma(X) = C(X)C^\top(X)$, $C(X)$ is a lower-triangular matrix. If $C(X)$ is

$$C(X) = \begin{pmatrix} c_{10} + c_{11}X & 0 \\ c_{20} + c_{21}X & c_{30} + c_{31}X \end{pmatrix},$$

the resulting correlation between Y_1 and Y_2 is $(c_{10} + c_{11}X)(c_{20} + c_{21}X)/[|(c_{10} + c_{11}X)|\sqrt{\{(c_{20} + c_{21}X)^2 + (c_{30} + c_{31}X)^2\}}]$. This clearly indicates that the effect of X on the correlation cannot be expressed in terms of a single parameter. There is not a single regression parameter to assess the effect of X on the correlation. All components of the parameter vector $c = (c_{10}, c_{11}, c_{20}, c_{21}, c_{30}, c_{31})^\top$ become highly coupled, making it challenging to discern their individual influence on the estimated quantities.

In the context of modeling covariance in terms of predictors, Chiu et al. (1996) proposed to model each entry of the logarithm of the covariance matrix as a linear combination of the predictor variables, referred as log-linear covariance models. This modeling technique preserves the positive definiteness of the covariance matrix (Stephenson (1965), Chap-

ter. 3). Although correlations can be obtained from the variance–covariance matrix, the model parameters are not directly interpretable in the context of correlations. Hoff and Niu (2012) used a parsimonious quadratic function of predictor variables to model the covariance matrix. For a large and sparse covariance matrix, Zou et al. (2017) proposed to write the covariance matrix as a linear combination of similarity matrices that are known functions of the predictors. The parameters (used in the linear combination) are estimated by maximizing the multivariate normal likelihood function with the restriction that all the eigenvalues of the covariance matrix are positive. These are excellent approaches if covariance modeling is of utmost importance.

We directly model the pairwise correlations and standard deviation in terms of the predictors. To bypass the critical but unsettling implementation of the positive definiteness constraint, we propose to work with the composite likelihood function stemming from the likelihood of pairwise response variables. The model parameters are then estimated by maximizing the logarithm of the composite likelihood.

Another challenge with this correlation model is its computational complexity, which tends to escalate with more predictors or traits. For instance, with q traits and p predictors, each of the q standard deviation models involves $q(1 + p)$ parameters, while there are $q(q - 1)/2(1 + p)$ parameters for $\binom{q}{2}$ pairwise correlations. Thus, the total parameter count is $(1+p)q(1+q)/2$, linear in predictor count and quadratic in trait count. To address this, we introduce a minorize–maximize (MM) algorithm (Hunter and Lange 2004) for parametric estimation. This algorithm decomposes the objective function into functions dependent on single parameters, significantly improving computational efficiency. While the MM algorithm has been successful in various fields (Huang et al. 2019; Zhou et al. 2019; Hunter and Runze 2005), its application in correlation modeling is innovative. Its effectiveness relies on identifying an appropriate minorizing function, crucial for optimizing complex objective functions like the logarithm of pairwise composite likelihood. Developing such a function lacks a standardized procedure and relies heavily on problem-specific considerations, demanding creative application of mathematical inequalities.

Numerical studies show that the use of the MM algorithm can reduce the computation time manifold compared to generic optimization routines. We have developed an `R` function, called CMPLE, standing for Correlation Modeling under Pairwise Likelihood Estimation, which can be readily applied while modeling correlations between multiple responses in terms of predictors (both continuous and categorical). Specifically, the methodological novelty lies in (1) direct modeling of correlations and standard deviations in terms of predictors and (2) MM-assisted optimization of the pairwise composite likelihood to obtain computationally efficient estimates.

The outline of the rest of the article is given below. Section 2 contains models, notations, and assumptions. Section 3 contains estimation and inferential techniques. Simulation studies are placed in Sect. 4, while analyses of the motivating dataset are in Sect. 5. Mathematical proofs and detailed derivations are collected in the supplementary materials.

2. MODELS AND NOTATIONS

2.1. BACKGROUND

To set the models and notations, assume that the observed data are collected from n independent units/subjects. For each unit, q traits (phenotypes) and p features (candidate genes, treatments, other predictors) are observed. Let $Y_{i,j}$ and $X_{i,r}$ be the jth trait and the rth feature of the ith unit, $j = 1, \ldots, q$, $r = 1, \ldots, p$, and $i = 1, \ldots, n$. The goal is to study the correlation between any pair of phenotypes and investigate the relation between the correlation and a set of features. Let us assume that conditional on the covariate $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,p})^\top$, $\boldsymbol{Y}_i = (Y_{i,1}, \ldots, Y_{i,q})^\top$ follows a multivariate normal distribution with mean $\boldsymbol{\mu}_i = 0$, and variance–covariance matrix $\boldsymbol{\Sigma}_i$. The variance–covariance matrix can be presented as $\boldsymbol{\Sigma}_i = \mathrm{Diag}(\sigma_{i,1}, \ldots, \sigma_{i,q})\, \boldsymbol{R}_i\, \mathrm{Diag}(\sigma_{i,1}, \ldots, \sigma_{i,q})$, where $\boldsymbol{R}_i = ((\rho_{i,j,k}))$ is the $q \times q$ correlation matrix for the q phenotypes from the ith subject, and the variance of $Y_{i,j}$ is denoted by $\sigma^2_{i,j}$.

2.2. CORRELATION MODELING

To achieve that goal, $\rho_{i,j,k}$, the pairwise correlation between $Y_{i,j}$ and $Y_{i,k}$, is written as $\rho_{i,j,k} = g^{-1}(\eta_{i,j,k})$, where $g : (-1, 1) \to (-\infty, \infty)$ is a known link function to transform the correlation to the linear predictor defined as $\eta_{i,j,k} = \delta_{j,k,0} + \sum_{r=1}^{p} \delta_{j,k,r} X_{i,r}$, where $\boldsymbol{\delta}_{j,k} = (\delta_{j,k,0}, \delta_{j,k,1}, \ldots, \delta_{j,k,p})^\top$ is the regression parameter. Observe that $\eta_{i,j,k} = g(\rho_{i,j,k})$, and we require that g to be a one-to-one function. There are many popular choices for the link function g. For the convenience, we take $g(\bullet) = \log\{(1 + \bullet)/(1 - \bullet)\}$. This results in

$$\rho_{i,j,k} = g^{-1}(\eta_{i,j,k}) = 1 - \frac{2}{1 + \exp(\eta_{i,j,k})} = 1 - \frac{2}{1 + \exp(\delta_{j,k,0} + \sum_{r=1}^{p} \delta_{j,k,r} X_{i,r})}. \tag{1}$$

The regression coefficient $\delta_{j,k,r}$ has a monotone linear relation with the correlation. Hence, we can interpret a predictor's effect on the correlation via the regression parameters $\delta_{j,k,r}$. Specifically, if $\delta_{j,k,r} > 0$ ($\delta_{j,k,r} < 0$), then the correlation between the jth and kth phenotype increases (decreases) with the rth feature, while other features remain unchanged.

2.3. STANDARD DEVIATION MODELING

The log-linear function is used to model the standard deviation of the phenotypes in terms of the features. Specifically, for the jth response (trait) and the ith experimental unit, the standard deviation is modeled as

$$\log(\sigma_{i,j}) = \alpha_{j,0} + \sum_{r=1}^{p} \alpha_{j,r} X_{i,r}. \tag{2}$$

The α parameters measure the effect of the features on the standard deviation. The proposed models for standard deviation and correlation can also be used to measure the average marginal effect (AME) of features on the correlations and standard deviations (see the supplementary material A.6).

3. ESTIMATION METHODOLOGY

3.1. COMPOSITE LIKELIHOOD

As mentioned previously, in our pairwise modeling, there is no guarantee that the correlation matrix R_i is positive definite. Thus, the model parameters cannot be estimated by maximizing the multivariate normal density function. With this, we propose to estimate the model parameters via the pairwise composite likelihood method. Now define $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \ \boldsymbol{\delta}^\top)^\top$, where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \ldots, \boldsymbol{\alpha}_q^\top)^\top$ and $\boldsymbol{\delta} = (\boldsymbol{\delta}_{1,2}^\top, \boldsymbol{\delta}_{1,3}^\top, \ldots, \boldsymbol{\delta}_{q-1,q}^\top)^\top$. The pairwise composite likelihood function for q responses is

$$CL(\boldsymbol{\theta}) = \prod_{j=1}^{q-1} \prod_{k=j+1}^{q} \mathcal{L}_{j,k}(\boldsymbol{\theta}), \tag{3}$$

where $\mathcal{L}_{j,k}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(Y_{i,j}, Y_{i,k}|\boldsymbol{X}_i)$ denotes the pairwise likelihood function for the jth and kth responses, and

$$f(Y_{i,j}, Y_{i,k}|\boldsymbol{X}_i) = \frac{1}{2\pi\sigma_{i,j}\sigma_{i,k}\sqrt{1-\rho_{i,j,k}^2}} \times \exp\left\{ -\frac{1}{2(1-\rho_{i,j,k}^2)}\left(\frac{Y_{i,j}^2}{\sigma_{i,j}^2} - \frac{2\rho_{i,j,k}Y_{i,j}Y_{i,k}}{\sigma_{i,j}\sigma_{i,k}} + \frac{Y_{i,k}^2}{\sigma_{i,k}^2}\right)\right\}.$$

The likelihood $\mathcal{L}_{j,k}$ refers to the joint distribution of the jth and kth traits, which is a function of the standard deviation of the jth trait, the standard deviation of the kth trait, and their pairwise correlation. So, even when the product in Equation (3) is over $k > j$, the composite likelihood is a function of all standard deviations and pairwise correlations.

The estimator of $\boldsymbol{\theta}$ is defined as $\widehat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta})$, where $\ell(\boldsymbol{\theta}) = \log\{CL(\boldsymbol{\theta})\}$. Note that the length of the $\boldsymbol{\theta}$-vector is $n_\theta = q \times (p+1) + \binom{q}{2} \times (p+1) = (p+1)q(q+1)/2$. This means the number of variables is quadratic only in the number of phenotypes and linear in the number of predictors. For a scenario with two features ($p = 2$) and four phenotypes ($q = 4$), n_θ is 30. For the scenario of $p = 6$ and $q = 4$, n_θ is 70. Thus, applying the standard Newton–Raphson method or its variant is very time-consuming as it will require repeated inversion of a large matrix. Therefore, we develop an MM algorithm which is computationally more efficient than direct maximization of $\ell(\boldsymbol{\theta})$ using the Newton–Raphson method.

3.2. THE MM ALGORITHM

The MM algorithm squarely depends on finding a suitable minorization function for the log of the composite likelihood, $\ell(\theta)$. Note that $\ell(\boldsymbol{\theta}) = \sum_{j<k} \ell_{j,k}(\boldsymbol{\alpha}, \boldsymbol{\delta})$, where $\ell_{j,k}(\boldsymbol{\alpha}, \boldsymbol{\delta})$

is the logarithm of the pairwise likelihood function

$$\ell_{j,k}(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{i=1}^{n}\Bigg\{\log(\sigma_{i,j}^2) + \log(\sigma_{i,k}^2) + \log(1-\rho_{i,j,k}^2) + \frac{1}{(1-\rho_{i,j,k}^2)}\left(\frac{Y_{i,j}^2}{\sigma_{i,j}^2} - \frac{2\rho_{i,j,k}Y_{i,j}Y_{i,k}}{\sigma_{i,j}\sigma_{i,k}} + \frac{Y_{i,k}^2}{\sigma_{i,k}^2}\right)\Bigg\}.$$

Before we state the main result whose proof is given in Appendix A.2, we introduce some notations. Define $D_{i,j,k}^{(t)} = \sigma_{i,j}^{(t)}\sigma_{i,k}^{(t)}(1-\rho_{i,j,k}^{(t)^2})$, $F_{i,j,k}^{(t)} = \sigma_{i,j}^{(t)}\sigma_{i,k}^{(t)}(1+\rho_{i,j,k}^{(t)})$, where $\sigma_{i,j}^{(t)}(\rho_{i,j,k})$ denotes $\sigma_{i,j}(\rho_{i,j,k}^{(t)})$ when the associated parameter $\boldsymbol{\alpha}(\boldsymbol{\delta})$ is replaced by its updated value at the tth iteration.

Theorem 1. *For any $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ in the parameter space,*

$$\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)}) = \sum_{j=1}^{p} g_1(\boldsymbol{\alpha}_j|\boldsymbol{\theta}^{(0)}) + \sum_{j<k}\sum g_2(\boldsymbol{\delta}_{j,k}|\boldsymbol{\theta}^{(0)}) + g_3(\boldsymbol{\theta}^{(0)})$$

is a minorization function of $\ell(\boldsymbol{\theta})$ such that and $\ell(\boldsymbol{\theta}) \geq \ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)})$ $\forall\boldsymbol{\theta}, \boldsymbol{\theta}_0$ and $\ell(\boldsymbol{\theta}) = \ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)})$ when $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$, where $g_1(\boldsymbol{\alpha}_j|\boldsymbol{\theta}^{(0)}) = \sum_{s:s<j}\psi_{1,s,j}(\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_j^{(0)}, j|\boldsymbol{\theta}^{(0)}) + \sum_{s:s>j}\psi_{1,j,s}(\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_j^{(0)}, j|\boldsymbol{\theta}^{(0)})$ for $j = 1,\ldots,q$, $g_2(\boldsymbol{\delta}_{j,k}|\boldsymbol{\theta}^{(0)}) = \psi_{2,j,k}(\rho_{j,k}|\boldsymbol{\theta}^{(0)})$ for $j \neq k$, and $g_3(\boldsymbol{\theta}^{(0)}) = \sum\sum_{j<k}\psi_{3,j,k}(\boldsymbol{\theta}^{(0)})$, with

$$\begin{aligned}
\psi_{1,j,k}(\boldsymbol{\alpha}_r - \boldsymbol{\alpha}_r^{(0)}, r|\boldsymbol{\theta}^{(0)}) = \sum_{i=1}^{n}\Bigg[&\left\{1 + \frac{(Y_{i,j}+Y_{i,k})^2}{2D_{i,j,k}^{(0)}} + \frac{(Y_{i,j}^2+Y_{i,k}^2)}{2F_{i,j,k}^{(0)}}\right\}\boldsymbol{Z}_i^\top(\boldsymbol{\alpha}_r^{(0)} - \boldsymbol{\alpha}_r) \\
&- \frac{Y_{i,r}^2}{4\sigma_{i,r}^{(0)^2}(1-\rho_{i,j,k}^{(0)^2})}\exp\{4\boldsymbol{Z}_i^\top(\boldsymbol{\alpha}_r^{(0)} - \boldsymbol{\alpha}_r)\} \\
&- \left\{\frac{(Y_{i,j}^2+Y_{i,k}^2)}{6D_{i,j,k}^{(0)}} + \frac{(Y_{i,j}+Y_{i,k})^2}{6F_{i,j,k}^{(0)}}\right\}\exp\{3\boldsymbol{Z}_i^\top(\boldsymbol{\alpha}_r^{(0)} - \boldsymbol{\alpha}_r)\}\Bigg],
\end{aligned}$$

$$\begin{aligned}
\psi_{2,j,k}(\rho_{j,k}|\boldsymbol{\theta}^{(0)}) = \sum_{i=1}^{n}\Bigg[&-\frac{1}{2}\log(1-\rho_{i,j,k}^2) - \left\{\frac{Y_{i,j}^2}{4\sigma_{i,j}^{(0)^2}(1-\rho_{i,j,k}^{(0)^2})} + \frac{Y_{i,k}^2}{4\sigma_{i,k}^{(0)^2}(1-\rho_{i,j,k}^{(0)^2})}\right\} \\
&\times\left(\frac{1-\rho_{i,j,k}^{(0)^2}}{1-\rho_{i,j,k}^2}\right)^2 + \frac{(Y_{i,j}+Y_{i,k})^2}{2D_{i,j,k}^{(0)}}\log\left(\frac{1-\rho_{i,j,k}^{(0)^2}}{1-\rho_{i,j,k}^2}\right) \\
&- \frac{(Y_{i,j}^2+Y_{i,k}^2)}{6D_{i,j,k}^{(0)}}\left(\frac{1-\rho_{i,j,k}^{(0)^2}}{1-\rho_{i,j,k}^2}\right)^3 - \frac{(Y_{i,j}+Y_{i,k})^2}{6F_{i,j,k}^{(0)}}\left(\frac{1+\rho_{i,j,k}^{(0)}}{1+\rho_{i,j,k}}\right)^3 \\
&+ \frac{(Y_{i,j}^2+Y_{i,k}^2)}{2F_{i,j,k}^{(0)}}\log\left(\frac{1+\rho_{i,j,k}^{(0)}}{1+\rho_{i,j,k}}\right)\Bigg],
\end{aligned}$$

$\boldsymbol{Z}_i = (1\ \boldsymbol{X}_i^\top)^\top$, *and the constant term* $\psi_{3,j,k}(\boldsymbol{\theta}^{(0)}) = \sum_{i=1}^n \{(Y_{i,j}+Y_{i,k})^2/2D_{i,j,k}^{(0)} + (Y_{i,j}^2 + Y_{i,k}^2)/2F_{i,j,k}^{(0)} - 0.5\log(\sigma_{i,j}^{(0)^2}\sigma_{i,k}^{(0)^2})\}$.

In the MM algorithm, we maximize the minorizing function ℓ^* rather than ℓ. The minorizing function ℓ^* is expressed as a summation of $g_1(\boldsymbol{\alpha}_1|\boldsymbol{\theta}^{(0)}), \ldots, g_1(\boldsymbol{\alpha}_p|\boldsymbol{\theta}^{(0)})$, and $g_2(\boldsymbol{\delta}_{1,2}|\boldsymbol{\theta}^{(0)}), \ldots, g_2(\boldsymbol{\delta}_{p-1,p}|\boldsymbol{\theta}^{(0)})$, this results in the separation of the parameters, which has a great advantage when optimizing a function with respect to a high-dimensional argument ($\boldsymbol{\theta}$ in our case). The functions g_1, g_2 are all differentiable functions and satisfy standard regularity conditions, and further being used in updating the parameter values. The parameter estimates are obtained by the gradient MM algorithm (Hunter and Lange 2004). Let $\boldsymbol{\theta}^{(t)}$ be the parameter value at the tth iteration, then $\boldsymbol{\theta}^{(t+1)}$ is obtained by one step Newton–Raphson method

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \left\{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})\right\}^{-1}_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}}\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}}. \tag{4}$$

The above step is repeated until the estimate converges. Specifically, we stop the above iteration when $\mathbf{1}^\top(|\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)}|/|\boldsymbol{\theta}^{(t)}|) < \epsilon_0$, a prespecified small number.

Note that the minorization function $\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ helps reducing the time complexity of the proposed method. For example, $\partial\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\alpha}_j = \partial g_1(\boldsymbol{\alpha}_j|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\alpha}_j$ is a function of $\boldsymbol{\alpha}_j$ only, and $\partial\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\delta}_{j,k} = \partial g_2(\boldsymbol{\delta}_{j,k}|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\delta}_{j,k}$ is a function of $\boldsymbol{\delta}_{j,k}$ only. Consequently, $\partial^2\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ is a block-diagonal matrix, and each block is a matrix of order $(p+1)\times(p+1)$ and this greatly enhances computational efficiency.

Specifically, the complexity of the inversion of each block matrix is in the order $O((p+1)^3)$. Thus, the complexity of one update of the MM algorithm is $O(nn_\theta + n(p+1)^2q(q+1)/2 + (p+1)^3q(q+1)/2)$. In other words, the complexity is $O(nn_\theta + n(p+1)n_\theta + (p+1)^2n_\theta)$, where $n_\theta = (1+p)q(1+q)/2$. On the other hand, the complexity of a direct optimization of $\ell(\boldsymbol{\theta})$ using the Newton–Raphson method is $O(nn_\theta + nn_\theta^2 + n_\theta^3)$. Alternatively, the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm can be used to optimize $\ell(\boldsymbol{\theta})$. This algorithm avoids large matrix inversion, so it has a square order complexity. Although the proportionality constant of the complexity order is unknown, the order of this complexity for BFGS is still larger than the complexity of the MM algorithm as long as $(p+1) < q(q+1)/2$, and this holds for our real data and the simulation scenarios.

The terms of Eq. (4) are $[\partial\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\theta}]_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}} = [\partial\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\alpha}_1, \ldots, \partial\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\alpha}_q, \partial\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\delta}_{1,2}, \ldots, \partial\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\delta}_{q-1,q}]_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}}$, and

$$\begin{aligned}\left(\frac{\partial\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})}{\partial\boldsymbol{\alpha}_j}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}} &= \sum_{s:s<j}\sum_{i=1}^n\left\{-1 + \frac{Y_{i,j}^2}{\sigma_{i,j}^{(t)^2}(1-\rho_{i,s,j}^{(t)^2})} - \frac{Y_{i,j}Y_{i,s}}{D_{i,j,s}^{(t)}} + \frac{Y_{i,j}Y_{i,s}}{F_{i,j,s}^{(t)}}\right\}\boldsymbol{Z}_i \\ &\quad + \sum_{s:s>j}\sum_{i=1}^n\left\{-1 + \frac{Y_{i,j}^2}{\sigma_{i,j}^{(t)^2}(1-\rho_{i,j,s}^{(t)^2})} - \frac{Y_{i,j}Y_{i,s}}{D_{i,j,s}^{(t)}} + \frac{Y_{i,j}Y_{i,s}}{F_{i,j,s}^{(t)}}\right\}\boldsymbol{Z}_i, \\ &\quad \text{for } j = 1, \ldots, q,\end{aligned}$$

$$\left(\frac{\partial \ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})}{\partial \boldsymbol{\delta}_{j,k}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}} = \sum_{i=1}^{n}\left\{\frac{\rho_{i,j,k}^{(t)}}{1-\rho_{i,j,k}^{(t)2}} - \frac{Y_{i,j}^2\rho_{i,j,k}^{(t)}}{\sigma^{(t)2}_{i,j}(1-\rho_{i,j,k}^{(t)2})^2} - \frac{Y_{i,k}^2\rho_{i,j,k}^{(t)}}{\sigma^{(t)2}_{i,k}(1-\rho_{i,j,k}^{(t)2})^2} + \frac{2Y_{i,j}Y_{i,k}\rho_{i,j,k}^{(t)}}{D_{i,j,s}^{(t)}(1-\rho_{i,j,k}^{(t)2})} + \frac{Y_{i,j}Y_{i,k}}{F_{i,j,s}^{(t)}(1+\rho_{i,j,k}^{(t)})}\right\}\frac{(1-\rho_{i,j,k}^{(t)2})}{2}\boldsymbol{Z}_i,$$

for $j < k = 1, \ldots, q$ and $\boldsymbol{Z}_i = (1\ \boldsymbol{X}_i^\top)^\top$. Furthermore, let $\boldsymbol{A} = \{\partial^2\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\}_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}}$, so $\boldsymbol{A} = \text{Diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_q, \boldsymbol{A}_{1,2}^\dagger, \ldots, \boldsymbol{A}_{q-1,q}^\dagger)$, with $\boldsymbol{A}_j = [\partial^2\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\alpha}_j\partial\boldsymbol{\alpha}_j^\top]_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}}$, $\boldsymbol{A}_{j,k}^\dagger = [\partial^2\ell^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})/\partial\boldsymbol{\delta}_{j,k}\partial\boldsymbol{\delta}_{j,k}^\top]_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}}$, and

$$\boldsymbol{A}_j = -\sum_{i=1}^{n}\Bigg(\sum_{s:s<j}\left[\frac{4Y_{i,j}^2}{\sigma_{i,j}^{(t)2}(1-\rho_{i,s,j}^{(t)2})} + \frac{3}{2\sigma_{i,j}^{(t)}\sigma_{i,s}^{(t)}}\left\{\frac{(Y_{i,j}^2+Y_{i,s}^2)}{(1-\rho_{i,s,j}^{(t)2})} + \frac{(Y_{i,j}+Y_{i,s})^2}{(1+\rho_{i,s,j}^{(t)})}\right\}\right] + \sum_{s:s<j}\left[\frac{4Y_{i,j}^2}{\sigma_{i,j}^{(t)2}(1-\rho_{i,j,s}^{(t)2})} + \frac{3}{2\sigma_{i,j}^{(t)}\sigma_{i,s}^{(t)}}\left\{\frac{(Y_{i,j}^2+Y_{i,s}^2)}{(1-\rho_{i,j,s}^{(t)2})} + \frac{(Y_{i,j}+Y_{i,s})^2}{(1+\rho_{i,j,s}^{(t)})}\right\}\right]\Bigg)\boldsymbol{Z}_i\boldsymbol{Z}_i^\top,$$

$$\begin{aligned}\boldsymbol{A}_{j,k}^\dagger = \sum_{i=1}^{n}\Bigg[&\frac{1+\rho_{i,j,k}^{(t)2}}{(1-\rho_{i,j,k}^{(t)2})^2} - \left(\frac{Y_{i,j}^2}{\sigma_{i,j}^{(t)2}} + \frac{Y_{i,k}^2}{\sigma_{i,k}^{(t)2}}\right)\times\frac{1+5\rho_{i,j,k}^{(t)2}}{(1-\rho_{i,j,k}^{(t)2})^3} \\ &+ \frac{1}{\sigma_{i,j}^{(t)}\sigma_{i,k}^{(t)}(1-\rho_{i,j,k}^{(t)2})^3}\left\{(Y_{i,j}+Y_{i,k})^2(1+\rho_{i,j,k}^{(t)2}) - (Y_{i,j}^2+Y_{i,k}^2)(1+7\rho_{i,j,k}^{(t)2})\right\} \\ &+ \frac{Y_{i,j}^2+Y_{i,k}^2-4(Y_{i,j}+Y_{i,k})^2}{2\sigma_{i,j}^{(t)}\sigma_{i,k}^{(t)}(1+\rho_{i,j,k}^{(t)})^3}\Bigg]\frac{(1-\rho_{i,j,k}^{(t)2})^2}{4}\boldsymbol{Z}_i\boldsymbol{Z}_i^\top \\ &+ \sum_{i=1}^{n}\Bigg[\frac{\rho_{i,j,k}^{(t)}}{1-\rho_{i,j,k}^{(t)2}} - \left(\frac{Y_{i,j}^2}{\sigma^{(t)2}_{i,j}} + \frac{Y_{i,k}^2}{\sigma^{(t)2}_{i,k}}\right)\times\frac{\rho_{i,j,k}^{(t)}}{(1-\rho_{i,j,k}^{(t)2})^2} - \frac{Y_{i,j}Y_{i,k}}{\sigma_{i,j}^{(t)}\sigma_{i,k}^{(t)}} \\ &\times\left\{\frac{2\rho_{i,j,k}^{(t)}}{(1-\rho_{i,j,k}^{(t)2})^2} + \frac{1}{(1+\rho_{i,j,k}^{(t)})^2}\right\}\Bigg]\frac{\rho_{i,j,k}^{(t)}(1-\rho_{i,j,k}^{(t)2})}{2}\boldsymbol{Z}_i\boldsymbol{Z}_i^\top.\end{aligned}$$

The above expressions are derived in Appendix A.3.

4. INFERENCE

Let $\boldsymbol{\theta}_0$ be the true parameter lies in an open subset of multidimensional Euclidean space. Assume that all predictors are in a compact subset of multidimensional Euclidean space, $\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i^\top$ has full rank, and other regularity conditions hold. Then following the standard asymptotic results (Varin et al. 2011), we obtain $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \longrightarrow N_{n_\theta}(\boldsymbol{0}, \boldsymbol{G}^{-1})$ in distribution, where N_{n_θ} denotes the n_θ-variate multivariate normal distribution. The Godambe information $\boldsymbol{G}$ is $\boldsymbol{G}(\boldsymbol{\theta}_0) = \boldsymbol{H}(\boldsymbol{\theta}_0)\boldsymbol{J}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^\top(\boldsymbol{\theta}_0)$, where $\boldsymbol{H}(\boldsymbol{\theta}) = E[-\nabla_{\boldsymbol{\theta}}\boldsymbol{U}(\boldsymbol{\theta}; D)]$, $\boldsymbol{J}(\boldsymbol{\theta}) = \text{var}[\boldsymbol{U}(\boldsymbol{\theta}; D)]$, where D denotes the data from randomly chosen subject or experimental unit, and $\boldsymbol{U}(\boldsymbol{\theta}; D)$ denotes the score function corresponding to this randomly

chosen subject. The information $\boldsymbol{G}(\boldsymbol{\theta}_0)$ is consistently estimated by $\widehat{\boldsymbol{H}}\widehat{\boldsymbol{J}}^{-1}\widehat{\boldsymbol{H}}^{\top}$, where $\widehat{\boldsymbol{H}} = -(1/n)\sum_{i=1}^{n}\nabla_{\boldsymbol{\theta}}\boldsymbol{U}(\boldsymbol{\theta}; D_i)|_{\widehat{\boldsymbol{\theta}}}$, $\widehat{\boldsymbol{J}} = (1/n)\sum_{i=1}^{n}\boldsymbol{U}(\boldsymbol{\theta}; D_i)\boldsymbol{U}(\boldsymbol{\theta}; D_i)^{\top}|_{\widehat{\boldsymbol{\theta}}}$, and

$$\begin{aligned}\boldsymbol{U}(\boldsymbol{\theta}; D_i) =& \frac{\partial}{\partial\boldsymbol{\theta}}\sum_{j}\sum_{j<k}\log\{f(Y_{i,j}, Y_{i,k}|\boldsymbol{X}_i)\}\\ =& -\frac{\partial}{2\partial\boldsymbol{\theta}}\Bigg[\sum_{j=1}^{q}(q-1)\log(\sigma_{i,j}^2) + \sum_{j<k}\sum\log(1-\rho_{i,j,k}^2)\\ &+\sum_{j<k}\sum\frac{1}{(1-\rho_{i,j,k}^2)}\left(\frac{Y_{i,j}^2}{\sigma_{i,j}^2} - \frac{2\rho_{i,j,k}Y_{i,j}Y_{i,k}}{\sigma_{i,j}\sigma_{i,k}} + \frac{Y_{i,k}^2}{\sigma_{i,k}^2}\right)\Bigg],\end{aligned}$$

and detailed expressions of $\boldsymbol{H}$ and $\boldsymbol{J}$ are given in Appendix A.4. Thus, the standard error of the rth component of $\widehat{\boldsymbol{\theta}}$ is the square root of the rth diagonal element of the inverse of $\widehat{\boldsymbol{H}}\widehat{\boldsymbol{J}}^{-1}\widehat{\boldsymbol{H}}^{\top}$. This standard error helps compute the Wald confidence interval for the parameter and can also be used for hypothesis testing.

5. SIMULATION STUDIES

5.1. SIMULATION DESIGN

Four different scenarios were considered. Recall that p and q denote the number of predictors and the number of phenotypes, respectively. For all scenarios, we set $q = 4$. For scenarios 1, 2, 3, and 4, we set $p = 2, n = 261$, $p = 2, n = 600$, $p = 6, n = 500$, and $p = 10, n = 1000$, respectively. The number of parameters are 30, 30, 70, and 110 for the four scenarios, respectively. Some of these configurations were chosen by closely following real datasets (Hoh et al. 2021). Each dataset contained information on X and Y from n independent units. The predictor X had p components, and each component was independently simulated from the Bernoulli(0.5) distribution. Next, $\boldsymbol{Y}$ was generated from $N_4(0, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i = \text{Diag}(\sigma_{i,1}, \ldots, \sigma_{i,q})\boldsymbol{R}_i\text{Diag}(\sigma_{i,1}, \ldots, \sigma_{i,q})$, where $\boldsymbol{R}_i = ((\rho_{i,j,k}))$. The true values of the parameter θ are given in the simulation tables.

5.2. METHOD OF ANALYSIS

For each scenario $\rho_{i,j,k}$'s and $\sigma_{i,j}$'s were modeled according to Eqs. (1) and (2), respectively. Under each scenario, 500 datasets were generated. Each dataset was analyzed by two approaches, 1) the proposed MM algorithm, and 2) the direct method where parameter estimates were obtained by directly maximizing the log-composite likelihood function. Under approach two, we used the `optim` function of `R` and chose to optimize using the L-BFGS-B method, a variant of the Broyden–Fletcher–Goldfarb–Shanno algorithm.

For both the approaches, the initial values for θ parameters were randomly generated from Normal(0, 0.10). Since our proposed method is an iterative optimization, we used the sum of the absolute relative difference between the parameter estimates in subsequent iterations to be less than 0.001 as the stopping criteria for the convergence for both simulation and real data analyses.

5.3. RESULTS

To save space results for scenarios 1, 2, and 4 are presented in Tables S1, S2 and S3 of Supplementary Materials. Results for scenario 3 are presented in Tables 1 and 2. We present the bias, the standard deviation of the estimates (SD), the estimated standard error (SE), the empirical coverage probability of the 95% Wald's confidence intervals, and the root-mean-squared error (RMSE) of the estimates for the proposed MM algorithm. All entries but the true parameter values in the tables are multiplied by 100. For example, if $\widehat{\theta}_i$ and SE_i denote the estimate and standard error for a generic parameter θ from the ith dataset, then the bias and SE presented in the table are $100(\sum_{i=1}^{500} \widehat{\theta}_i/500 - \theta)$ and $100 \sum_{i=1}^{500} SE_i/500$, respectively. Also, the direct optimization method results are qualitatively similar to the MM algorithm. Hence, they are not presented in the tables.

The key points are: (1) Parameter estimate biases are negligible across various sample sizes and different p values, (2) standard errors (SEs) closely match standard deviations (SDs), indicating accurate capture of asymptotic estimator variability, and (3) the empirical coverage probabilities are pretty close to 0.95. All of these indicate that the method of estimation works well, and asymptotic properties of the estimator hold. Both SD and SE decrease with sample size (see Tables S1 and S2 in Supplementary Materials). Even with moderate (Tables 1 and 2) or large parameter counts (Table S3), the MM algorithm performs exceptionally well. Notably, biases and SDs (including RMSE) are notably larger for δ parameters compared to α parameters, suggesting greater uncertainty (less information) in correlation parameters than standard deviations. To address a reviewer's concern, we checked if the reconstructed variance–covariance matrix is positive definite. Using simulation scenario 4, we obtained the distribution of the smallest eigenvalue of the true variance–covariance matrix for every given value of the covariate (ground truth). On the other hand, after estimating the parameters concerning the pairwise correlation and standard deviations for the

Table 1. Results of α parameters from the simulation study for scenario 3 with $n = 500$, $p = 6$, $q = 4$

Par	True	Bias	SD	SE	CP	RMSE	Par	True	Bias	SD	SE	CP	RMSE
$\alpha_{1,0}$	−1.0	−0.8	8.7	8.0	92.4	8.8	$\alpha_{3,0}$	1.0	−1.7	8.7	8.1	92.6	8.8
$\alpha_{1,1}$	1.0	0.5	6.0	6.0	94.2	6.0	$\alpha_{3,1}$	0.3	0.4	6.5	6.1	92.4	6.5
$\alpha_{1,2}$	0.2	0.2	6.1	6.0	94.8	6.1	$\alpha_{3,2}$	0.1	−0.3	6.5	6.1	93.8	6.5
$\alpha_{1,3}$	−0.4	−0.4	6.2	6.0	95.2	6.2	$\alpha_{3,3}$	0	0.3	6.4	6.1	93.2	6.4
$\alpha_{1,4}$	0.3	−0.1	6.0	6.0	95.4	6.0	$\alpha_{3,4}$	−0.2	0.1	6.6	6.1	92.6	6.6
$\alpha_{1,5}$	0	0.5	6.2	6.0	93.4	6.2	$\alpha_{3,5}$	0.1	0.5	6.5	6.1	93.8	6.5
$\alpha_{1,6}$	−0.5	−0.1	6.5	6.0	93.8	6.5	$\alpha_{3,6}$	−0.2	0.6	6.3	6.1	93.8	6.3
$\alpha_{2,0}$	2.0	−0.4	8.4	8.1	93.4	8.4	$\alpha_{4,0}$	−1.0	−0.7	8.4	8.0	92.8	8.4
$\alpha_{2,1}$	0.2	−0.1	6.3	6.0	93.8	6.2	$\alpha_{4,1}$	−0.5	−0.2	6.5	6.0	92.0	6.5
$\alpha_{2,2}$	−0.5	0	6.2	6.0	94.6	6.2	$\alpha_{4,2}$	0	0	6.6	6.1	91.2	6.6
$\alpha_{2,3}$	−0.4	−0.3	6.7	6.2	92.8	6.7	$\alpha_{4,3}$	0.3	−0.6	6.3	6.0	93.4	6.3
$\alpha_{2,4}$	0	0	6.4	6.0	93.8	6.4	$\alpha_{4,4}$	−0.2	0.3	6.2	6.0	93.6	6.2
$\alpha_{2,5}$	0.3	−0.5	6.1	6.0	93.4	6.1	$\alpha_{4,5}$	−0.2	0.1	6.1	6.0	95.0	6.1
$\alpha_{2,6}$	0	−0.2	6.2	6.1	93.8	6.2	$\alpha_{4,6}$	0.2	0.2	6.2	6.0	94.4	6.2

All entries except for the true parameter values of the table are multiplied by 100. Par: Parameter, SD: standard deviation, SE: standard error, CP: 95% coverage probability, RMSE: root-mean-squared error

Table 2. Results of δ parameters from the simulation study for scenario 3 with $n = 500$, $p = 6$, $q = 4$

Par	True	Bias	SD	SE	CP	RMSE	Par	True	Bias	SD	SE	CP	RMSE
$\delta_{1,2,0}$	0.2	1.2	23.5	23.1	94.8	23.5	$\delta_{2,3,0}$	0	1.1	24.5	23.5	92.8	24.5
$\delta_{1,2,1}$	0.5	−2.4	17.7	17.3	95.6	17.8	$\delta_{2,3,1}$	0	0.3	19.3	17.8	93.4	19.3
$\delta_{1,2,2}$	0	−4.4	17.2	17.2	94.6	17.8	$\delta_{2,3,2}$	−0.2	3.1	18.5	17.9	93.6	18.7
$\delta_{1,2,3}$	−0.7	3.3	18.3	17.4	94.4	18.6	$\delta_{2,3,3}$	0	−2.5	18.9	17.8	93.8	19.0
$\delta_{1,2,4}$	0.3	−2.7	16.9	17.2	95.2	17.1	$\delta_{2,3,4}$	0	1.7	17.0	17.8	95.0	17.1
$\delta_{1,2,5}$	0	1.8	17.6	17.2	93.0	17.7	$\delta_{2,3,5}$	0	−1.1	17.8	17.8	94.4	17.8
$\delta_{1,2,6}$	−0.8	3.0	18.0	17.5	93.4	18.2	$\delta_{2,3,6}$	0	−2.6	18.4	17.8	93.0	18.6
$\delta_{1,3,0}$	0.5	−3.4	25.0	23.4	93.0	25.2	$\delta_{2,4,0}$	0.3	1.7	23.3	23.0	94.4	23.4
$\delta_{1,3,1}$	0.2	−0.5	17.9	17.6	94.2	17.9	$\delta_{2,4,1}$	0	2.7	18.6	17.2	93.0	18.8
$\delta_{1,3,2}$	0.5	−1.4	17.7	17.7	95.0	17.8	$\delta_{2,4,2}$	−1.0	4.1	18.1	17.6	93.6	18.6
$\delta_{1,3,3}$	0	1.9	19.3	17.5	92.0	19.3	$\delta_{2,4,3}$	0.4	−4.5	18.0	17.2	93.2	18.5
$\delta_{1,3,4}$	0.2	−3.9	19.2	17.5	92.4	19.6	$\delta_{2,4,4}$	−0.3	1.4	17.0	17.2	94.4	17.1
$\delta_{1,3,5}$	−0.5	5.4	18.8	17.7	93.6	19.5	$\delta_{2,4,5}$	0.5	−1.4	17.6	17.2	93.6	17.7
$\delta_{1,3,6}$	0	1.0	18.4	17.5	93.8	18.4	$\delta_{2,4,6}$	0.3	−6.0	18.4	17.2	92.4	19.3
$\delta_{1,4,0}$	1.0	−1.4	26.8	23.5	90.8	26.8	$\delta_{3,4,0}$	−1.0	0.5	24.4	23.4	92.6	24.4
$\delta_{1,4,1}$	0.2	−2.1	19.6	17.7	92.0	19.7	$\delta_{3,4,1}$	0.2	1.3	18.3	17.6	95.2	18.3
$\delta_{1,4,2}$	−0.5	−0.1	18.1	17.8	94.8	18.1	$\delta_{3,4,2}$	−0.2	2.2	16.9	17.7	95.4	17.0
$\delta_{1,4,3}$	−0.2	0.1	18.9	17.7	93.8	18.9	$\delta_{3,4,3}$	−0.1	1.0	18.5	17.6	94.0	18.5
$\delta_{1,4,4}$	0.5	−2.8	18.3	17.8	94.2	18.5	$\delta_{3,4,4}$	0.2	2.9	17.8	17.7	95.2	18.0
$\delta_{1,4,5}$	0.1	4.6	18.5	17.7	92.6	19.1	$\delta_{3,4,5}$	0.6	−2.8	17.9	17.7	95.2	18.1
$\delta_{1,4,6}$	−0.1	−0.9	18.0	17.7	95.0	18.0	$\delta_{3,4,6}$	−0.1	−1.4	18.6	17.6	92.8	18.6

All entries except for the true parameter values of the table are multiplied by 100. Par: Parameter, SD: standard deviation, SE: standard error, CP: 95% coverage probability, RMSE: root-mean-squared error

Table 3. Average computation time (in seconds) to analyze a single dataset using the MM algorithm and direct optimization (DOP) via the optim function with the "L-BFGS-B" method under different simulation scenarios

	Simulation scenario			
	1	2	3	4
MM	50.21	163.94	335.76	522.42
DOP	237.82	569.70	2040.36	13486.20

phenotypes, we have reconstructed the variance–covariance matrix and computed the minimum eigenvalue for every covariate value. These steps were replicated under 100 datasets. Figure 3 of the supplementary material shows the two distributions. While their centers align, the estimated distribution can contain negative values due to our method's lack of guarantee for positive definiteness.

The proposed method has a clear advantage in terms of computation time over the direct optimization method (using a form of gradient descent). The proposed method is much faster than the direct method (see Table 3 for details). The time gain increases with the number of parameters. For instance, the computation time using the direct optimization was more than four days, while MM took about a couple of hours only, for the entire simulation under scenario 4. All simulations were done on a 2.8Ghz Intel Xeon E5-1603 processor. Appendix A.5 contains results on some further numerical experiment; these results suggest

the computation time increases linearly and polynomially with the number of parameters for the proposed method and the direct optimization of the composite likelihood, respectively.

6. DATA EXAMPLE

6.1. SPECIFIC BACKGROUND FOR THE PHOTOSYNTHETIC EXPERIMENT

Photosynthesis in plant chloroplasts involves a series of light-driven reactions. Light energy is absorbed by chlorophyll molecules, transferring energy to the photochemical reaction centers, photosystem I (PSI) and photosystem II (PSII), where electrons are excited, resulting in electron transfer reactions through complex mechanisms. Under ideal conditions, photosynthesis is highly efficient and has near-unity quantum efficiency. Yet, under many real-world conditions, light energy input exceeds the capacity of the chloroplast to process it, leading to the buildup of reactive intermediates and subsequent photodamage (Kramer and Evans 2011; Murchie et al. 2015). Chloroplasts can protect themselves by dissipating the excess light energy through non-photochemical quenching (NPQ), which safely dissipates light energy as heat. NPQ can prevent the buildup of reactive intermediates but at the cost of energy loss. NPQ exists in two main forms: qE and qI. qE serves as the primary photoprotective mechanism, while qI acts as a slower backup process under severe environmental stress. Although both contribute to NPQ, they differ significantly in biological origin and dependence on other processes. Understanding the fractions of NPQ attributable to qE and qI is crucial, as they serve as physiological indicators of a plant's adaptation to varying environmental conditions.

In the current work, we estimate "total" NPQ designated NPQt, which represents the sum of and does not distinguish between qE, qI and other NPQ forms. However, we expect the for the qE mechanism to see a positive between NPQt and the extent of thylakoid pmf, which we can estimate using an independently measured parameter, ECSt. This predicted association should be modified or broken down when qE is replaced with qI. For example, under conditions or in mutants where the qE response is inhibited, photodamage or photoinhibition should ensue, increasing qI. Such cases are illustrated in Appendix A.1. In these cases, the qI will be associated with a loss of photochemical activity, in turn leading to decreased electron and proton transfer, ultimately resulting in decreased pmf. These effects should result in a negative correlation between NPQt in the form of qI and ECSt, i.e., the opposite of that expected for qE.

Figure 1 illustrates varying associations (positive and negative) between NPQt and ECSt under different environments and QTL markers. These findings suggest that qE and qI responses are influenced by both environmental factors and genetics. Understanding how these correlation patterns coincide with genetic markers and diverse environmental conditions can unveil crucial components governing the transition between primary and fallback behaviors, offering insights for enhancing crop productivity.

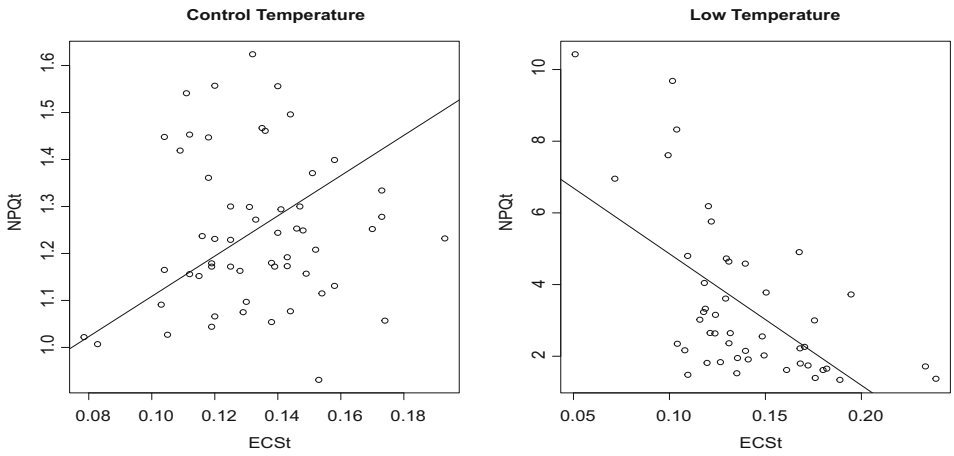

Figure 1. Real data example where we identify examples of photoprotection and photodamage being regulated by chromosomes 4 and 9 at different environmental conditions .

6.2. PLANT MATERIALS

A population of 90 cowpeas (*Vigna unguiculata (L.) Walp.*) recombinant inbred lines (RILs) comprising two parent lines and 88 offspring were utilized in this study. These offspring were generated from a cross between the two parent lines, the tolerant cultivar California Blackeye 27 (CB27), developed by the University of California, Riverside, and a sensitive breeding line, 24-125B-1, developed by the Institute de Recherche Agricole pour le Développement (IRAD, Cameroon). Our experimental involved 360 plants. Of these, 180 plants comprised two parent lines and 88 offspring, replicated twice across two separate batches, and were subjected to the control temperature (CT, 29°C/19°C (day/night)). The remaining 180 plants were exposed to the low temperature (LT, chilling stress 19°C/13°C (day/night)). For both treatments, two seeds were sown in a 3.5 × 3.5-inch pot filled with Suremix potting medium (Michigan Grower Products Inc, USA) and watered with half-strength Hoagland's nutrient solution. Plants were arranged in a reach-in growth chamber (BioChambers, Winnipeg, Canada) in a Randomized Complete Block Design (RCBD) in two batches, with each batch including two replicates of each RILs. Data for both batches were pooled together for the final analysis.

Photosynthetic responses were measured using MultispeQ 2.0 hand-held instruments (Kuhlgert et al. 2016). We used the following $q = 4$ phenotypes as responses, namely (1) steady-state PS II quantum yields, ϕ_{II}, (2) non-photochemical quenching, NPQt, (3) Q_A redox state PS II center opened, qL, and (4) thylakoid pmf (proton motive force), ECSt. On the genotyping aspect, a total of 1141 single nucleotide polymorphism (SNP) markers of genotype data of *CB27 ×24-125B-1* were obtained from Lonardi et al. (2019). Plants of the RIL population are homozygous for each marker in the two parental lines indicated by either AA (having the allele from *CB27*, the tolerant, maternal line), or BB (having the allele from *24-125B-1*, the sensitive, paternal line), which are further dummy coded as binary (0, 1) features, where 0 (1) stands for the AA (BB) allele at a given marker locus.

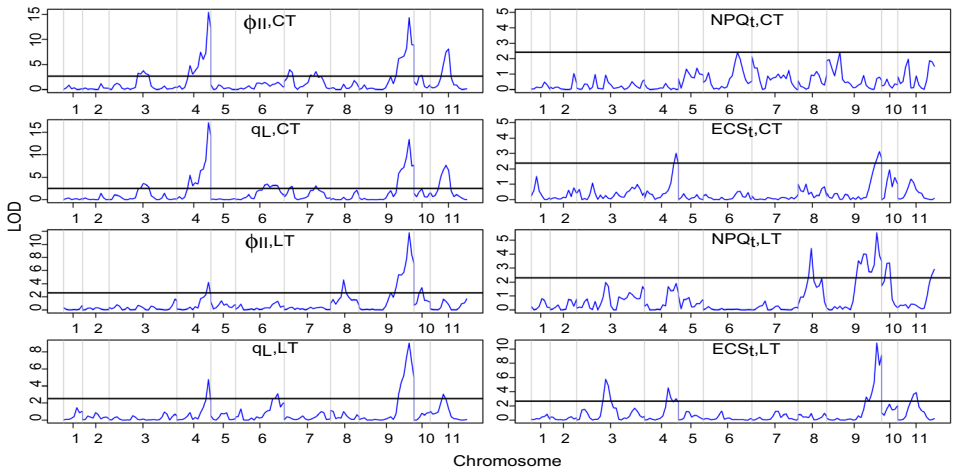

Figure 2. QTL plot of different phenotypes used in the Cowpea RIL data. LOD threshold for each phenotype is marked by the bold horizontal line. QTL with a LOD higher than that can be considered significant. Chromosomes are marked by the vertical lines .

6.3. DATA PREPROCESSING AND QTL MAPPING

Before conducting the CMPLE analyses, adjustments were made to individual phenotypes to account for design effects by fitting linear mixed-effects models. Let $P_{i,l,m}$ represent a phenotype corresponding to the ith genotype, lth batch, and the mth replication, then it was modeled as $P_{i,l,m} = \mu + \gamma_l + \tau_{m(l)} + e_{i,l,m}$, $i = 1, \ldots, n$, $l = 1, \ldots, L$, $m = 1, \ldots, M$, where μ, γ_l, $\tau_{m(l)}$, and e denote the overall mean, the random effect due to the lth batch, the nested random effect due to the m replication within the lth batch, and the residual error, respectively. The R package `nlme` was used for model fitting under the assumption that $\gamma_l \sim \text{Norma}(0, \sigma^2_\gamma)$, $\tau_{m(l)} \sim \text{Norma}(0, \sigma^2_\tau)$, and the residuals were utilized for further analysis. Particularly, the preprocessed response was calculated as $(LM)^{-1} \sum_{l=1}^{L} \sum_{m=1}^{M} (P_{i,l,m} - \widehat{\mu} - \widehat{\gamma_l} - \widehat{\tau}_{m(l)})$ for the phenotype in the downstream analyses, where $\widehat{\mu}$, $\widehat{\gamma_l}$ and $\widehat{\tau}_{m(l)}$ are the estimates of the model components. Note that the downstream analysis relies on the assumption that the preprocessed response is independently distributed. This assumption constitutes an approximation due to shared estimated effects between different response values.

To narrow down model covariates to candidate genes, we employed two methods to select specific SNPs affecting the four phenotypes. First, we conducted individual QTL analysis (Fig. 2) using the Multiple QTL Mapping (MQM) model via the `Rqtl` package (Broman and Sen 2009). LOD thresholds were determined through permutation analysis (`mqmpermutation` and `mqmscan` functions in R), with 1000 permutations and a p value cutoff of 0.05. This revealed two candidate loci on chromosomes 4 (59.64 cm) and 9 (86.93 cm) as significant SNPs under both conditions.

Next, we used a complementary GWAS analysis, where each of the 1141 markers was tested for each phenotype one at a time while adjusting for the environment. Significant markers were chosen based on the Benjamini–Hochberg false discovery rate (FDR) $\leq$ 0.0005. The number of significant markers for the four phenotypes, ϕ_{II}, NPQt, qL, and

ECSt, was 105, 23, 85, and 51, respectively. Finally, a set of 12 markers, common from the individual sets, were selected for further investigation. Our main aim here was to use CMPLE to identify markers most likely to predict the phenotypic association across different environments.

6.4. METHOD OF ANALYSES

We split our data analysis into two parts. First, we assessed if candidate loci on chromosomes 4 and 9 could classify phenotypic associations as synergistic or antagonistic using two markers (G1 and G2) and an environmental variable (E) to analyze (1) main effects and (2) main effects + two-way interactions, e.g., G1×E, G2×E, and G1×G2. Next, we aimed to uncover genetic effects and gene–environment interactions on cross-phenotype association with 12 markers identified through GWAS. Similar to single-variant analysis in GWAS, we modeled phenotype standard deviation and pairwise correlations with predictors (one marker at a time), environment, and their interaction.

For the first framework, we fit the following model to $\sigma_{\phi_{II}}$, the standard deviation of phenotype ϕ_{II}, in terms of the predictors, $\sigma_{\phi_{II}} = \exp(\alpha_{1,0} + \alpha_{1,1}\text{G1} + \alpha_{1,2}\text{G2} + \alpha_{1,3}\text{E})$. Framework two and three were then followed by: $\sigma_{\phi_{II}} = \exp(\alpha_{1,0} + \alpha_{1,1}\text{G1} + \alpha_{1,2}\text{G2} + \alpha_{1,3}\text{E} + \alpha_{1,4}\text{G1} \times \text{G2} + \alpha_{1,5}\text{G1} \times \text{E} + \alpha_{1,6}\text{G2} \times \text{E})$, and $\sigma_{\phi_{II}} = \exp(\alpha_{1,0} + \alpha_{1,1}\text{Marker} + \alpha_{1,2}\text{E} + \alpha_{1,3}\text{Marker} \times \text{E})$

A similar model was fit to the standard deviation of the other phenotypes. Simultaneously, we fit the following model to the pairwise correlations between phenotypes ϕ_{II} and NPQt, $\rho_{\phi_{II}\&\text{NPQt}} = 1 - 2/\{1 + \exp(\delta_{1,2,0} + \delta_{1,2,1}\text{G1} + \delta_{1,2,2}\text{G2} + \delta_{1,2,3}\text{E})\}$. For framework two, we model $\rho_{\phi_{II}\&\text{NPQt}} = 1 - 2/\{1 + \exp(\delta_{1,2,0} + \delta_{1,2,1}\text{G1} + \delta_{1,2,2}\text{G2} + \delta_{1,2,3}\text{E} + \delta_{1,2,4}\text{G1} \times \text{G2} + \delta_{1,2,5}\text{G1} \times \text{E} + \delta_{1,2,6}\text{G2} \times \text{E})\}$ and for framework three, we have $\rho_{\phi_{II}\&\text{NPQt}} = 1 - 2/\{1 + \exp(\delta_{1,2,0} + \delta_{1,2,1}\text{Marker} + \delta_{1,2,2}\text{E} + \delta_{1,2,3}\text{Marker} \times \text{E})\}$. Similarly, the remaining five pairwise correlations, $\rho_{\phi_{II}\&qL}$, $\rho_{\phi_{II}\&\text{ECSt}}$, $\rho_{\text{NPQt}\&qL}$, $\rho_{\text{NPQt}\&\text{ECSt}}$, $\rho_{qL\&\text{ECSt}}$ were modeled in terms of the predictors.

Before applying the MM algorithm, we subtracted the respective mean from the four phenotypes to have a mean of zero. For the above-described models, there were a total of 40 parameters including $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$. Here, $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\delta}^\top)^\top$. We set the initial value of each component of $\boldsymbol{\theta}$ to the random numbers generated from Normal(0, 0.15).

6.5. INTERPRETATION

We infer cross-phenotype association based on the estimated $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ parameters. We estimated the model parameters for the first set of analyses with two candidate markers and calculated the AIC values to determine which fits the data best. The AIC score of the main effects model (number of estimated parameters= 40) came out to be 61.59, compared to 121.77 for the main + two-way interaction effects model (number of estimated parameters= 70). Thus, based on the lower AIC, we are interpreting the results (Tables 4 and 5 for $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$) from the main effects model only. The results from the main + two-way interaction effects model can be found in Table S4 and Table S4 in Supplementary Materials.

Table 4. Parameter estimates and the 95% confidence interval in parentheses of the parameters of the standard deviation model for the measured phenotypes from the cowpea dataset

	ϕ_{II}	NPQ_t	qL	ECS_t
Intercept	−2.61 (−2.73, −2.49)	−0.37 (−0.51, −0.23)	−2.17 (−2.29, −2.05)	−3.18 (−3.43, −2.93)
Marker 1	−0.04 (−0.16, 0.08)	−0.43 (−0.61, −0.25)	0.03 (−0.13, 0.19)	0.02 (−0.16, 0.20)
Marker 2	0.08 (−0.04, 0.20)	0.45 (0.29, 0.61)	−0.07 (−0.21, 0.07)	−0.24 (−0.42, −0.06)
Environment	0.19 (0.07, 0.31)	0.45 (0.29, 0.61)	−0.14 (−0.30, 0.02)	0.16 (0.01, 0.32)

Table 5. Parameter estimates and 95% confidence interval in parentheses of the parameters of pairwise correlation among the measured phenotypes from the cowpea dataset

	Intercept	Marker 1	Marker 2	Environment
ϕ_{II}&NPQ_t	−1.35 (−1.72, −0.98)	−0.45 (−0.82, −0.08)	0.12 (−0.25, 0.49)	−0.71 (−1.10, −0.32)
ϕ_{II}&qL	1.82 (1.49, 2.15)	0.29 (−0.02, 0.60)	−0.32 (−0.63, −0.01)	−0.30 (−0.63, 0.03)
ϕ_{II}&ECS_t	0.16 (−0.23, 0.55)	−0.65 (−1.02, −0.28)	0.38 (−0.01, 0.77)	0.17 (−0.18, 0.52)
NPQ_t&qL	0.35 (−0.12, 0.82)	−0.54 (−0.95, −0.13)	0.23 (−0.18, 0.64)	−0.42 (−0.75, −0.09)
NPQ_t&ECS_t	0.82 (0.45, 1.19)	0.13 (−0.24, 0.50)	−0.14 (−0.55, 0.27)	−1.05 (−1.44, −0.66)
qL&ECS_t	1.01 (0.58, 1.44)	−0.52 (−0.95, −0.09)	−0.01 (−0.42, 0.40)	−0.47 (−0.88, −0.06)

The $\boldsymbol{\alpha}_j$-parameters (j is the index for phenotype) can be viewed as the conditional effect of each predictor variable on the standard deviation of jth phenotype. For example, the conditional effect of G1 on the standard deviation of NPQt was estimated to be −0.43, with a standard error (SE) of 0.08. This means that in our population, if there is a change from allele AA to allele BB at G1, the conditional standard deviation of NPQt is estimated to be decreased by 35%, while all other predictors remain unchanged. Similarly, in Table 5, we collect $\boldsymbol{\delta}_{j,k}$ estimates that can be used to calculate the conditional effect of each predictor on the correlation between (j, k) phenotype pair. For example, the estimated regression parameter of G1 on the pairwise correlation of ϕ_{II} and NPQt was −0.45 (SE = 0.18). This means in our population, if the G1 allele changes from AA to BB, the conditional pairwise correlation between ϕ_{II} and NPQt is estimated to be decreased by 0.13 (using Eq. 1) when G2 is at allele AA and the E is set to CT. Also, after appropriate transformations, the intercept terms represent the baseline conditional standard deviation and pairwise correlation among phenotypes when all the markers are fixed at allele AA and E is fixed at CT. For example, in Table 5, the estimated intercept −1.35 (SE = 0.19) under the column ϕ_{II}&NPQt implies that the estimated correlation between ϕ_{II} and NPQt is $1 - 2/\{1 + \exp(-1.35)\} = -0.59$ when all the predictors are at their baseline.

6.6. BIOLOGICAL IMPLICATIONS

We have identified cross-phenotype associations by assessing the direction of genetic loci across various environments, classifying them as synergistic or antagonistic. If predictors influence two phenotypes similarly, it is deemed synergistic; if they act oppositely, it is antagonistic. Tables 4 and 5 show coefficients' signs, e.g., G1's positive effect and G2's

Table 6. Pairwise correlation estimates and 95% confidence interval in parentheses of the measured phenotypes from all genetic combinations of Marker 1 and Marker 2 from the cowpea dataset

E		Genotype *AAAA*	*AABB*	*BBAA*	*BBBB*
CT	ϕ_{II}&NPQ$_t$	−0.59 (−0.70, −0.45)	−0.55 (−0.68, −0.37)	−0.72 (−0.78, −0.64)	−0.69 (−0.78, −0.57)
	ϕ_{II}&qL	0.72 (0.63, 0.79)	0.64 (0.53, 0.72)	0.78 (0.73, 0.83)	0.71 (0.64, 0.78)
	ϕ_{II}&ECS$_t$	0.08 (−0.11, 0.27)	0.27 (0.08, 0.43)	−0.24 (−0.38, −0.09)	−0.06 (−0.26, 0.15)
	NPQ$_t$&qL	0.17 (−0.06, 0.39)	0.28 (0.13, 0.42)	−0.10 (−0.23, 0.04)	0.02 (−0.19, 0.23)
	NPQ$_t$&ECS$_t$	0.39 (0.22, 0.54)	0.33 (0.11, 0.52)	0.44 (0.34, 0.54)	0.38 (0.18, 0.55)
	qL&ECS$_t$	0.47 (0.29, 0.62)	0.46 (0.33, 0.58)	0.24 (0.04, 0.42)	0.23 (0.01, 0.43)
LT	ϕ_{II}&NPQ$_t$	−0.77 (−0.84, −0.68)	−0.75 (−0.81, −0.67)	−0.85 (−0.90, −0.78)	−0.83 (−0.88, −0.77)
	ϕ_{II}&qL	0.64 (0.53, 0.73)	0.54 (0.40, 0.65)	0.72 (0.63, 0.79)	0.63 (0.51, 0.73)
	ϕ_{II}&ECS$_t$	0.16 (−0.04, 0.36)	0.34 (0.18, 0.49)	−0.16 (−0.31, 0.03)	0.03 (−0.16, 0.21)
	NPQ$_t$&qL	−0.04 (−0.25, 0.18)	0.03 (−0.12, 0.18)	−0.30 (−0.42, −0.16)	−0.19 (−0.38, −0.02)
	NPQ$_t$&ECS$_t$	−0.19 (−0.30, −0.09)	−0.18 (−0.31, −0.05)	−0.05 (−0.24, 0.15)	−0.03 (−0.25, 0.19)
	qL&ECS$_t$	0.26 (0.04, 0.46)	0.26 (0.04, 0.45)	0.01 (−0.15, 0.16)	0.01 (−0.23, 0.23)

E≡ Environment, CT: control temperature, LT: low temperature

negative effect on the correlation between ϕ_{II} and qL, suggesting parallel regulatory pathways in photosynthesis. Table 6 presents estimated correlations and 95% CIs for all possible genetic-environmental combinations.

The pairwise correlations between NPQt and ECSt exhibited a positive trend under control temperature (CT), while a negative association prevailed under low temperature (LT). This suggests that genetic variations in chromosomes 4 and 9 predominantly influenced photochemistry via the qE mechanism under CT, and through the qI mechanism under LT, corroborating findings from Hoh et al. (2021) regarding additional adaptive mechanisms. Specifically under LT, combinations AAAA and AABB showed negative correlations between NPQt and ECSt, indicating a greater likelihood of chromosome 4 variations influencing the qI mechanism. Moreover, under LT, we observed different correlation patterns between NPQt and qL for various genetic groups. For AAAA and AABB combinations, they were uncorrelated, while for BBAA and BBBB combinations, they exhibited negative correlations. This suggests that when downstream processes are disrupted, electrons accumulate in Q_A, leading to a more reduced state of Q_A (resulting in decreased qL) and enhanced qI activity. With the accumulation of qI, the slope gradient between NPQt and qL gradually strengthens until reaching a point where the negative associations between NPQt and qL start to diminish. Interestingly, within AAAA and AABB combinations under LT, no correlation between NPQt and qL was observed. These findings were further utilized to estimate the markers' marginal effects on correlations; additional details can be found in Section A.6 of the Supplementary Materials.

In Table 7, using the 12 co-localized markers identified through single-variant analysis, we found four significant phenotype pairs (ϕ_{II} & NPQt, NPQt & qL, NPQt &ECSt, and qL &ECSt) for Gene × Environment interaction. Across all 12 markers, the effect of genes on the aforementioned pairwise interactions turned out to be positive (synergistic), whereas Gene × Environment interaction effect was negative (antagonistic). This suggests that the electron transfer chain linking these phenotypes can be stratified into different

Table 7. Parameter estimates of the pairwise correlation parameters based on the 12 co-localized markers selected through single-variant analysis

SNP		ϕ_{II}&NPQ$_t$		ϕ_{II}&qL		ϕ_{II}&ECS$_t$		NPQ$_t$&qL		NPQ$_t$&ECS$_t$		ECS$_t$&qL	
1	Int	−1.93	(−2.22, −1.64)	2.19	(1.90, 2.48)	−0.13	(−0.46, 0.20)	−0.43	(−0.74, −0.12)	0.60	(0.27, 0.93)	0.40	(0.05, 0.75)
	G	1.34	(0.91, 1.77)	−0.28	(−0.67, 0.11)	0.51	(−0.02, 1.04)	1.29	(0.84, 1.74)	0.60	(0.11, 1.09)	0.86	(0.37, 1.35)
	E	0.43	(−0.06, 0.92)	−0.22	(−0.69, 0.25)	−0.21	(−0.66, 0.24)	0.44	(0.01, 0.87)	0.28	(−0.28, 0.77)	−0.10	(−0.65, 0.45)
	G × E	−2.27	(−2.94, −1.60)	−0.23	(−0.94, 0.48)	0.47	(−0.26, 1.20)	−1.81	(−2.48, −1.14)	−2.39	(−3.08, −1.70)	−1.31	(−2.09, −0.53)
2	Int	−1.56	(−2.21, −1.64)	2.90	(1.91, 2.48)	−0.23	(−0.36, 0.30)	−0.45	(−0.64, −0.17)	0.64	(0.26, 0.88)	0.42	(0.10, 0.85)
	G	1.26	(0.94, 1.83)	−0.27	(−0.63, 0.16)	0.55	(−0.06, 0.98)	1.17	(0.79, 1.74)	0.61	(0.14, 1.11)	0.87	(0.38, 1.35)
	E	0.44	(−0.07, 0.89)	−0.17	(−0.66, 0.15)	−0.26	(−0.61, 0.31)	0.44	(0.02, 0.89)	0.24	(−0.21, 0.74)	−0.14	(−0.69, 0.48)
	G × E	−2.31	(−2.93, −1.64)	−0.26	(−0.91, 0.42)	0.44	(−0.23, 1.25)	−1.70	(−2.32, −1.03)	−2.43	(−3.21, −1.73)	−1.33	(−2.13, −0.48)
3	Int	−1.93	(−2.22, −1.64)	2.19	(1.90, 2.48)	−0.13	(−0.46, 0.20)	−0.43	(−0.74, −0.12)	0.60	(0.27, 0.93)	0.40	(0.05, 0.75)
	G	1.34	(0.91, 1.77)	−0.28	(−0.67, 0.11)	0.51	(−0.02, 1.04)	1.29	(0.84, 1.74)	0.60	(0.11, 1.09)	0.86	(0.37, 1.35)
	E	0.42	(−0.07, 0.91)	−0.22	(−0.69, 0.25)	−0.21	(−0.66, 0.24)	0.44	(0.01, 0.87)	0.28	(−0.21, 0.77)	−0.10	(−0.65, 0.45)
	G × E	−2.36	(−3.03, −1.69)	−0.22	(−0.93, 0.49)	0.47	(−0.26, 1.20)	−1.81	(−2.48, −1.14)	−2.40	(−3.09, −1.71)	−1.31	(−2.09, −0.53)
4	Int	−2.02	(−2.31, −1.73)	2.22	(1.93, 2.51)	−0.15	(−0.48, 0.18)	−0.50	(−0.81, −0.19)	0.59	(0.26, 0.92)	0.36	(0.01, 0.71)
	G	1.47	(1.04, 1.90)	−0.32	(−0.73, 0.09)	0.56	(0.03, 1.09)	1.42	(0.97, 1.87)	0.63	(0.14, 1.12)	0.95	(0.44, 1.46)
	E	0.54	(0.05, 1.03)	−0.22	(−0.71, 0.27)	−0.15	(−0.62, 0.32)	0.52	(0.09, 0.95)	0.25	(−0.22, 0.72)	−0.05	(−0.60, 0.50)
	G × E	−2.53	(−3.20, −1.86)	−0.20	(−0.93, 0.53)	0.39	(−0.35, 1.13)	−1.97	(−2.66, −1.28)	−2.43	(−3.14, −1.72)	−1.50	(−2.26, −0.74)
5	Int	−2.02	(−2.31, −1.73)	2.22	(1.93, 2.51)	−0.15	(−0.48, 0.18)	−0.50	(−0.81, −0.19)	0.59	(0.26, 0.92)	0.36	(0.01, 0.71)
	G	1.48	(1.05, 1.91)	−0.32	(−0.73, 0.09)	0.56	(0.03, 1.09)	1.42	(0.97, 1.87)	0.63	(0.14, 1.12)	0.95	(0.44, 1.46)
	E	0.54	(0.05, 1.03)	−0.22	(−0.71, 0.27)	−0.15	(−0.62, 0.32)	0.52	(0.09, 0.95)	0.25	(−0.22, 0.72)	−0.05	(−0.60, 0.50)
	G × E	−2.68	(−3.33, −2.03)	−0.19	(−0.92, 0.54)	0.39	(−0.35, 1.13)	−1.97	(−2.66, −1.28)	−2.47	(−3.18, −1.76)	−1.50	(−2.26, −0.74)
6	Int	−2.03	(−2.32, −1.74)	2.21	(1.92, 2.50)	−0.18	(−0.51, 0.15)	−0.50	(−0.81, −0.19)	0.61	(0.28, 0.94)	0.33	(−0.02, 0.68)
	G	1.54	(1.11, 1.97)	−0.30	(−0.71, 0.11)	0.67	(0.16, 1.18)	1.47	(1.02, 1.92)	0.60	(0.09, 1.11)	1.05	(0.56, 1.54)
	E	0.54	(0.05, 1.03)	−0.20	(−0.67, 0.27)	−0.15	(−0.60, 0.30)	0.51	(0.08, 0.94)	0.25	(−0.22, 0.72)	−0.04	(−0.59, 0.51)
	G × E	−2.54	(−3.21, −1.87)	−0.25	(−0.96, 0.46)	0.38	(−0.36, 1.12)	−1.99	(−2.68, −1.30)	−2.44	(−3.15, −1.73)	−1.54	(−2.28, −0.80)

Table 7. (*Continued*)

SNP		ϕ_{II}&NPQ$_t$		ϕ_{II}&qL		ϕ_{II}&ECS$_t$		NPQ$_t$&qL		NPQ$_t$&ECS$_t$		ECS$_t$&qL	
7	Int	−1.98	(−2.27, −1.69)	2.07	(1.78, 2.36)	−0.18	(−0.53, 0.17)	−0.38	(−0.69, −0.07)	0.64	(0.31, 0.97)	0.39	(0.02, 0.76)
	G	1.41	(0.98, 1.84)	−0.03	(−0.44, 0.38)	0.63	(0.12, 1.14)	1.16	(0.71, 1.61)	0.50	(0.01, 0.99)	0.85	(0.34, 1.36)
	E	0.51	(0.02, 1.00)	−0.05	(−0.54, 0.44)	−0.11	(−0.58, 0.36)	0.39	(−0.04, 0.82)	0.14	(−0.35, 0.63)	−0.10	(−0.67, 0.47)
	G × E	−2.41	(−3.06, −1.76)	−0.54	(−1.27, 0.19)	0.32	(−0.42, 1.06)	−1.69	(−2.36, −1.02)	−2.22	(−2.93, −1.51)	−1.34	(−2.10, −0.58)
8	Int	−1.93	(−2.24, −1.62)	2.01	(1.70, 2.32)	−0.09	(−0.44, 0.26)	−0.31	(−0.64, 0.02)	0.59	(0.24, 0.94)	0.50	(0.13, 0.87)
	G	1.26	(0.83, 1.69)	0.05	(−0.36, 0.46)	0.36	(−0.15, 0.87)	0.97	(0.52, 1.42)	0.60	(0.11, 1.09)	0.55	(0.04, 1.06)
	E	0.42	(−0.09, 0.93)	−0.06	(−0.55, 0.43)	−0.23	(−0.7, 0.24)	0.33	(−0.12, 0.78)	0.18	(−0.31, 0.67)	−0.23	(−0.82, 0.36)
	G × E	−2.16	(−2.83, −1.49)	−0.60	(−1.27, 0.07)	0.55	(−0.18, 1.28)	−1.44	(−2.09, −0.79)	−2.19	(−2.90, −1.48)	−0.87	(−1.65, −0.09)
9	Int	−2.01	(−2.30, −1.72)	2.22	(1.93, 2.51)	−0.13	(−0.46, 0.20)	−0.49	(−0.80, −0.18)	0.58	(0.25, 0.91)	0.37	(0.02, 0.72)
	G	1.43	(1.00, 1.86)	−0.33	(−0.74, 0.08)	0.51	(−0.02, 1.04)	1.38	(0.93, 1.83)	0.64	(0.15, 1.13)	0.90	(0.39, 1.41)
	E	0.52	(0.03, 1.01)	−0.25	(−0.74, 0.24)	−0.22	(−0.69, 0.25)	0.52	(0.09, 0.95)	0.32	(−0.17, 0.81)	−0.08	(−0.65, 0.49)
	G × E	−2.42	(−3.09, −1.75)	−0.16	(−0.87, 0.55)	0.49	(−0.24, 1.22)	−1.91	(−2.58, −1.24)	−2.47	(−3.16, −1.78)	−1.33	(−2.11, −0.55)
10	Int	−2.00	(−2.29, −1.71)	2.22	(1.93, 2.51)	−0.13	(−0.46, 0.20)	−0.49	(−0.80, −0.18)	0.58	(0.25, 0.91)	0.37	(0.02, 0.72)
	G	1.42	(0.99, 1.85)	−0.33	(−0.74, 0.08)	0.51	(−0.02, 1.04)	1.38	(0.93, 1.83)	0.64	(0.15, 1.13)	0.90	(0.39, 1.41)
	E	0.51	(0.02, 1.00)	−0.25	(−0.74, 0.24)	−0.22	(−0.69, 0.25)	0.52	(0.09, 0.95)	0.32	(−0.17, 0.81)	−0.08	(−0.65, 0.49)
	G × E	−2.37	(−3.04, −1.7)	−0.17	(−0.88, 0.54)	0.49	(−0.24, 1.22)	−1.91	(−2.58, −1.24)	−2.46	(−3.15, −1.77)	−1.33	(−2.11, −0.55)
11	Int	−1.91	(−2.20, −1.62)	2.03	(1.72, 2.34)	−0.05	(−0.38, 0.28)	−0.31	(−0.64, 0.02)	0.59	(0.24, 0.94)	0.56	(0.19, 0.93)
	G	1.20	(0.77, 1.63)	0.02	(−0.39, 0.43)	0.26	(−0.25, 0.77)	0.95	(0.50, 1.40)	0.57	(0.08, 1.06)	0.42	(−0.09, 0.93)
	E	0.29	(−0.26, 0.84)	−0.21	(−0.74, 0.32)	−0.09	(−0.56, 0.38)	0.30	(−0.15, 0.75)	−0.11	(−0.66, 0.44)	−0.16	(−0.75, 0.43)
	G × E	−1.94	(−2.63, −1.25)	−0.42	(−1.11, 0.27)	0.29	(−0.44, 1.02)	−1.36	(−2.01, −0.71)	−1.78	(−2.51, −1.05)	−1.06	(−1.86, −0.26)
12	Int	−1.91	(−2.20, −1.62)	2.03	(1.72, 2.34)	−0.05	(−0.38, 0.28)	−0.31	(−0.64, 0.02)	0.59	(0.24, 0.94)	0.56	(0.19, 0.93)
	G	1.20	(0.77, 1.63)	0.02	(−0.39, 0.43)	0.26	(−0.25, 0.77)	0.95	(0.50, 1.40)	0.57	(0.08, 1.06)	0.42	(−0.09, 0.93)
	E	0.29	(−0.26, 0.84)	−0.21	(−0.74, 0.32)	−0.09	(−0.56, 0.38)	0.30	(−0.15, 0.75)	−0.11	(−0.66, 0.44)	−0.16	(−0.75, 0.43)
	G × E	−1.96	(−2.65, −1.27)	−0.42	(−1.11, 0.27)	0.29	(−0.44, 1.02)	−1.36	(−2.01, −0.71)	−1.78	(−2.51, −1.05)	−1.06	(−1.86, −0.26)

The 95% confidence intervals are given in parentheses. Int: Intercept, G: Gene, E: Environment

models based on different G × E signals. Additionally, we observed the absolute values of the estimated correlation parameters between qL & ECSt vary significantly between markers. Specifically, for marker 8 (Chromosome 9, 93.15 cm), the estimated regression parameter for their correlation was −0.87 (SE = 0.4), the lowest among all 12 markers. This may indicate an alternative photosynthetic pathway governed by marker 8 under low temperatures, warranting further investigation.

7. DISCUSSION

We have developed the CMPLE workflow (see supplementary Figure 5) to bridge the gap in the PheWAS studies by exploiting the correlation structure among phenotypes based on genetic and environmental variables. The proposed methodology demonstrates that considering the cross-phenotype correlations across the G × E domain can identify novel biological processes and help recover pertinent information regarding different regulatory pathways associated with genetic variations.

We sparse out the genetic marker effects of specific SNPs in chromosomes 4 and 9, which under the control temperature favor photoprotective mechanisms, whereas, under low temperature, they are more consistent with modulation of qI. Also, under low-temperature conditions, where the photosynthetic machinery favors the qI mechanism, we have identified the correlations between NPQt and qL to be changing with specific genetic configurations, e.g., uncorrelated NPQt and qL (for combinations AAAA and AABB) and negatively correlated NPQt and qL (for combinations BBAA and BBBB). This provides evidence for a subsequent hypothesis that the gradient between NPQt and qL can further modulate the qI mechanism under low temperatures. This can be interpreted as one example of reflecting epistasis, where multiple genetic components interact in complex ways to further modulate regulatory pathways inside a system.

CMPLE can provide genome-enabled predictions for correlation between multiple traits by distinguishing traits that arise through distinct mechanisms under diverse environmental conditions and different development stages. This capability should allow breeders to target better combinations of traits that can complement or reinforce desired traits, especially those related to improved responses to other environmental conditions, most notably for application to climate-resilient agriculture.

It is essential to highlight that mean regression modeling and correlation modeling reveal different aspects of association in the phenotype–genotype space. Therefore, whether we work with the residual responses (residual obtained after regressing the phenotypes on the candidate genes) or the mean zero responses (obtained after subtracting the crude means from respective phenotypes), the results of the correlation analysis remain unchanged. The current approach was developed for a moderate number of predictors. A regularized estimation technique for a large number of predictors is a part of our future research. A limitation of our real data analysis is that we used preprocessed data, and the way preprocessing is typically done may result in correlated observations, leading to potentially biased parameter estimates. Alternatively, analyzing raw data and incorporating design-related parameters, such as effects due to replication or batch, could be considered. For instance, given the

raw data for the observed phenotype, $P_{i,j,l,m}$ for the jth trait from the mth replicate of the ith genotype under the lth batch, one could model the conditional mean ($\mu_{i,j,l,m}$), standard deviation ($\sigma_{i,j,l,m}$) and pairwise correlation ($\rho_{i,j,k,l,m}$) as follows: $\mu_{i,j,l,m} = \beta_{j,0} + \sum_{r=1}^{p} \beta_{j,r} X_{i,r} + \gamma_{j,l} + \tau_{j,m(l)}$, $\log(\sigma_{i,j,l,m}) = \alpha_{j,0} + \sum_{r=1}^{p} \alpha_{j,r} X_{i,r} + \gamma^{o}_{j,l} + \tau^{o}_{j,m(l)}$, $\eta_{i,j,k,l,m} = g(\rho_{i,j,k,l,m}) = \delta_{j,k,0} + \sum_{r=1}^{p} \delta_{j,k,r} X_{i,r} + \gamma^{\dagger}_{j,k,l} + \tau^{\dagger}_{j,k,m(l)}$. In this context, the parameters α, β, and δ represent the regression coefficients corresponding to the covariates, and one would assume that the two random vectors $(\gamma_{j,l}, \gamma^{o}_{j,l}, \gamma^{\dagger}_{j,l})^{\top}$ and $(\tau_{j,(m)l}, \tau^{o}_{j,m(l)}, \tau^{\dagger}_{j,m(l)})^{\top}$ follow two multivariate distributions. Additionally, for full generality, these random vectors should be allowed to exhibit dependence across different traits. The likelihood function incorporating such dependencies necessitates multidimensional integration with respect to these random vectors. However, computing such a likelihood is non-trivial, and a suitable method for analyzing data with numerous model parameters is yet to be developed and would be an interesting future research project.

In summary, this work represents a significant advance in modeling pairwise correlation and standard deviation in terms of predictor variables. The modeling is also accompanied by a novel estimation technique that boosts the optimization problem involving many parameters. Besides methodology development, we have shown that joint inference of standard deviations and correlations among phenotypes can be used to test the co-segregation of genetically resolved associations between different traits and improve the precision of phenotype network structure.

SUPPLEMENTARY MATERIALS

The supplementary material file contains proof of the theorem, a detailed derivation of expressions, further simulation results, and the CMPLE workflow. Our software implementation of CMPLE is available at https://github.com/ChattoAbhijnan/CMPLE, and we have incorporated a demo example in Appendix A.7. The data supporting this study's findings are available from the MSU-DOE Plant Research Laboratory, but restrictions apply to their availability. These data are not publicly available. The data are, however, available from the authors upon reasonable request and with the permission of MSU-DOE Plant Research Laboratory.

ACKNOWLEDGEMENTS

This work was supported by the DOE Office of Science, Basic Energy Sciences under awards DE-FG02-91ER20021, for the high throughput phenotyping tools, and DE-SC0007101 for detailed photosynthetic measurements. DMK was partially supported by MSU AgBioScience. TM was partially supported by DMS-2124605 and DMS-1924724. The authors would like to thank the entire review team for their constructive comments, which have improved the quality of the work. This article was written by the first author in his private capacity. No official support or endorsement by the NIH, National Institute of Environmental Health, is intended or should be inferred.

Declarations

Conflict of interest The authors declare no conflict of interest.

[Received December 2022. Revised April 2024. Accepted April 2024.]

REFERENCES

Broman KW, Sen S (2009) A guide to QTL mapping with R/qtl. Springer, New York

Chiu TYM, Leonard T, Tsui K-W (1996) The matrix-logarithmic covariance model. J Am Stat Assoc 91:198–210

Daniels MJ, Pourahmadi M (2002) Bayesian analysis of covariance matrices and dynamic models for longitudinal data. Biometrika 89:553–566

Gianola D, Fernando RL (2020) A multiple-trait Bayesian lasso for genome-enabled analysis and prediction of complex traits. Genetics 214:305–331

Hall MA, Verma A, Brown-Gentry KD, Goodloe R, Boston J, Wilson S, McClellan B, Sutcliffe NB et al (2014) Detection of pleiotropy through a phenome-wide association study of epidemiologic data as part of the environmental architecture for genes linked to environment (eagle) study. PLoS Genet 10:e1004678

Hoff PD, Niu X (2012) A covariance regression model. Stat Sin 22:729–753

Hoh D, Osei-Bonsu I, Chattopadhyay A, et al (2021) Genetic variation in photosynthetic responses to chilling modulates proton motive force, cyclic electron flow and photosystem ii photoinhibition. Authorea

Huang JZ, Liu N, Pourahmadi M, Liu L (2006) Covariance matrix selection and estimation via penalised normal likelihood. Biometrika 93:85–98

Huang X, Xu J, Tian G (2019) On profile MM algorithms for Gamma frailty survival models. Stat Sin 29:895–916

Hunter DR, Lange K (2004) A tutorial on MM algorithms. Am Stat 58:30–37

Hunter DR, Runze L (2005) Variable selection using MM algorithms. Ann Stat 33:1617–1642

Kanazawa A, Chattopadhyay A, Kuhlgert S, Tuitupou H, Maiti T, Kramer DM (2021) Light potentials of photosynthetic energy storage in the field: what limits the ability to use or dissipate rapidly increased light energy? R Soc Open Sci 8:211102

Kramer DM, Evans JR (2011) The importance of energy balance in improving photosynthetic productivity. Plant Physiol 155:70–78

Kramer DM, Johnson G, Kiirats O, Edwards GE (2004) New fluorescence parameters for the determination of qa redox state and excitation energy fluxes. Photosynth Res 79:209–218

Kuhlgert S, Austic G, Zegarac R, Osei-Bonsu I, Hoh D, Roth MG, Bi K, TerAvest D, Weebadde P, Kramer DM (2016) MultispeQ Beta: a tool for large-scale plant phenotyping connected to the open PhotosynQ network. R Soc Open Sci 3:160592

Lonardi S, Muñoz-Amatriaín M, Liang Q, Shu S, Wanamaker SI, Lo S, Tanskanen J, Schulman AH, Zhu T, Luo MC, Alhakami H, Ounit R (2019) The genome of cowpea (*Vigna unguiculata* [L.] Walp.). Plant J 98:767–782

Murchie EH, Ali A, Herman T (2015) Photoprotection as a trait for rice yield improvement: status and prospects. Rice 8:1–9

Pourahmadi M (1999) Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. Biometrika 86:677–690

Stephenson G (1965) An introduction to matrices, sets, and groups. Longmans, London

Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. Stat Sin 21:5–42

Zhou H, Hu L, Zhou J, Lange K (2019) MM algorithms for variance components models. J Comput Graph Stat 28:350–361

Zou T, Lan W, Wang H, Tsai C-L (2017) Covariance regression analysis. J Am Stat Assoc 112:266–281

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.